



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Evaluating an Expectation-Driven QUD Model of Discourse Interpretation

Citation for published version:

Kehler, A & Rohde, H 2017, 'Evaluating an Expectation-Driven QUD Model of Discourse Interpretation', *Discourse Processes*, vol. 54, no. 3, pp. 219-238. <https://doi.org/10.1080/0163853X.2016.1169069>

Digital Object Identifier (DOI):

[10.1080/0163853X.2016.1169069](https://doi.org/10.1080/0163853X.2016.1169069)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Discourse Processes

Publisher Rights Statement:

This is an Accepted Manuscript of an article published by Taylor & Francis in *Discourse Processes* on 4/4/2016, available online: <http://www.tandfonline.com/10.1080/0163853X.2016.1169069>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Evaluating an Expectation-Driven QUD Model of Discourse Interpretation

Andrew Kehler

University of California San Diego, Department of Linguistics, La Jolla, CA, USA

Hannah Rohde

University of Edinburgh, Department of Linguistics, Scotland, UK

Abstract

According to Question-Under-Discussion (QUD) models of discourse interpretation, clauses cohere with the preceding context by virtue of providing answers to (usually implicit) questions that are situated within a speaker's goal-driven strategy of inquiry. In this paper, we present four experiments that examine the predictions of a QUD model of interpretation when cast in terms of an integrated, expectation-driven model of discourse processing. The results of these studies together support the predictions of the model, demonstrating that contextual cues affect comprehenders' expectations about ensuing QUDs (Experiment 1), QUD expectations in turn influence the interpretation of discourse-dependent linguistic expressions (Experiment 2), and the biases associated with those expressions in turn influence the anticipation of QUDs (Experiments 3a/b).

Evaluating an Expectation-Driven QUD Model of Discourse Interpretation

In a conversation that recently transpired, a linguist named Andy said the following to his collaborator Hannah:

(1) We should write a paper. There's a special issue of Discourse Processes coming up that will be dedicated to QUDs.

A conversation ensued in which Hannah asked Andy about the timeline for the special issue, they compared schedules, and so forth: Hannah clearly inferred that the paper Andy had in mind would be *for* the special issue. That's strange; Andy didn't actually say that.

Of course, Hannah did what hearers typically do: They don't generally interpret co-occurring statements as independent facts about the world, but instead assume that the statements are somehow relevant to one another. According to contemporary theories of discourse interpretation, this assumption of relevance, and the inferences that comprehenders must draw to maintain it, result from the need to establish the COHERENCE of a discourse. In fitting with the theme of the special issue, the work reported here examines the Question Under Discussion (QUD) analysis of coherence, pursued in different instantiations by researchers including Carlson (1983), Klein and Stutterheim (1989), Kuppevelt (1995), Roberts (1996), Ginzburg (1996), Larsson (1998), and Ginzburg and Sag (2000).

In QUD analyses, discourses are structured by question/answer relationships: Roughly speaking, an utterance is coherent insofar as it provides an answer to a (generally implicit) question that is relevant to the preceding discourse. For instance, in unexceptional circumstances, the second sentence of (1) will be interpreted to answer the question *Why?* with respect to the eventuality described by the first. Importantly, the second sentence can only be understood to answer this question if it is assumed that the paper is being written for the special issue. Although by no means entailed by the passage, this assumption is easily accommodated by the addressee in the absence of any beliefs to the contrary.

Indeed, if Andy had later commented *Oh, I don't think we should submit to the special issue; I was thinking about a different paper*, Hannah would have been rightfully confused, and perhaps even have dropped her joint projects with Andy in favor of working with other, presumably more sane, collaborators.

QUD analyses reflect a complex interplay between the linguistic characteristics of a speaker's contribution and her underlying communicative intentions in producing it. In the QUD analysis set forth by Roberts (2012/1998), for instance, the coherence of discourse fundamentally depends on the conversational goals that interlocutors have and the STRATEGIES OF INQUIRY they employ to satisfy them. She suggests, following Stalnaker (1979), that discourse is to be viewed as an attempt by conversational participants to discover and share "the ways things are" (or, to phrase it another way, to answer the question *What is the way things are?*). Thus, by engaging in a conversation, the interlocutors agree to jointly adopt goals that center around finding answers to this question. This in turn will generally necessitate the adoption and satisfaction of subgoals centered on answering sub-questions, giving rise to a hierarchical discourse structure. On the QUD model, therefore, understanding a discourse requires that hearers not only understand the particular utterances in the discourse, but also recover the questions they answer, and situate those questions within the underlying strategy of inquiry.

This paper sets out to explore the ramifications of adopting a QUD model for the study of sentence and discourse processing, and conversely, to examine using sentence and discourse processing data how QUDs might be utilized during interpretation. Two possibilities for processing models come immediately to mind. First, according to a REACTIVE model of processing, comprehenders will first compute interpretations of individual utterances, and then use linguistic cues, world knowledge, and inference to determine the identity of the QUD and to integrate the utterance into the larger discourse. On this type of bottom-up view of language processing, the interpretation of linguistic input affects the identification of the QUD, but not the other way around. Whereas this picture is perfectly compatible with many linguistic theories that invoke QUDs in treatments of different phenomena, it seems unlikely in light of a substantial line of research that shows that interpretation is incremental and probabilistic at various levels of

linguistic processing (Crain & Steedman, 1985; Altmann, 1988; and many others). This research would instead suggest an alternative EXPECTATION-DRIVEN model of processing, according to which comprehenders are anticipators who use contextual cues to generate expectations about where the discourse is going in terms of the QUDs to which subsequent sentences will provide answers. On this model, evidence for the ensuing QUD may come from any of a number of contextual sources; examples that have been posited in the literature include focus marking and accent placement (Roberts, 2012/1998; Büring 2003), the semantic type of the verb appearing in a context sentence (Kehler et al., 2008), aspect (H. Rohde et al. 2006), the typicality of objects of transfer in transfer-of-possession events (H. Rohde et al., 2007, Experiment 1), and the existence of indefinites, disjunctions, and/or existential quantifiers in a context sentence (Anderbois, 2011), among others. The advantage of computing such expectancies for the comprehender is that the ensuing linguistic material will be easier to process when their expectations prove to be correct. Confirmation of the predictions of a QUD analysis would thus suggest that QUDs need to be taken into account in theories of discourse-dependent linguistic forms if these theories are to be empirically adequate.¹

This paper describes a series of experiments designed to examine three predictions of an expectation-driven QUD model of discourse processing. First, the model predicts that discourse participants will use contextual clues to create expectations about the ensuing QUD that a subsequent utterance will answer.² Experiment 1 asks whether we find such evidence. The results from participants producing questions in a simple dialog task demonstrate that different contexts yield different distributions of questions. The second prediction is that comprehenders will bring QUD-level expectations to bear when interpreting linguistic expressions that are highly discourse dependent.

¹ A remaining question, not addressed by the current work, is whether such evidence strongly favors QUD analyses over other analyses of discourse coherence on offer. We return to this question in the General Discussion.

² Von Kuppevelt (1995) in fact proposed that a set of possible questions arise in a listener's mind during discourse comprehension, and that these questions are entertained as ones that may be answered in subsequent discourse. He leaves open the question of which factors in a context contribute to the anticipation of these upcoming questions, however.

Experiment 2 examines third-person pronouns as a case study: Can we find evidence that QUDs affect the incremental assignment of referents to pronouns? The results answer this question affirmatively using a self-paced reading task in which the participants were told what QUD a pronoun-containing clause would answer. Lastly, a fully integrative expectation-driven model would also predict that the interaction between inferring implicit QUDs and interpreting linguistic expressions should be bidirectional, that is, that encountering linguistic material will introduce new biases that will in turn update expectations about QUDs. Experiments 3a/b address this question with passage completion studies, showing that the mere occurrence of an ambiguous pronoun in the completion prompt – a form which is known to carry its own referential biases – has a significant effect on the distribution of questions that the completions answer. The results are highly consistent with the predictions of an integrated, expectation-driven QUD model of discourse processing in which the anticipation of QUDs affects the interpretation of linguistic expressions, and the interpretation of linguistic expressions in turn affects the anticipation of QUDs.

Experiment 1: The Effect of Context on QUDs

According to the expectation-driven QUD model, addressees not only establish the implicit QUD that a clause answers when they interpret discourses, they also use contextual cues to generate expectations about ensuing QUDs. We investigated this prediction as part of a larger study that also examined the predictions of a RELATIONAL approach to discourse coherence (e.g., Hobbs 1990, Asher and Lascarides 2003). According to the relational approach, establishing coherence between utterances is based on making the inferences necessary to meet the constraints imposed by one of a set of context-bearing relevancy relations, or COHERENCE RELATIONS. In the case of example (1), the coherence relation that is most readily inferred is what various authors (e.g., Hobbs 1990, Asher and Lascarides 2003) have called EXPLANATION, according to which the second utterance explains (i.e., denotes a cause of or gives a reason for) the eventuality described by the first clause. In addition to Explanation, other commonly posited relations include Result, Occasion, Parallel, Elaboration, and Violated Expectation. Table 1 lists the set of coherence relations that we will refer to in this paper (definitions taken or adapted

from Hobbs 1990). The structures of larger discourses are then built recursively: Starting with the assumption that sentences are discourse segments, larger segments (all the way up to the entire discourse) result by relating two smaller segments by a coherence relation.

The larger study was a discourse completion experiment that utilized three distinct prompt types: two of which examined coherence relations in a monologue setting and one of which examined QUDs in a dialog setting. The monologue component of the experiment, the results of which were previously reported in Kehler et al. (2008), provided evidence that different contexts give rise to varying expectations about the coherence relations that are likely to follow. Contexts were varied between ones that employed implicit causality (IC) verbs (Garvey, Caramazza, & Yates, 1976; Caramazza, Grober, Garvey, & Yates, 1977; McKoon, Greene, & Ratcliff, 1993; Koornneef & Berkum, 2006, *inter alia*, henceforth IC) as in (2a) and (2c) with ones using non-IC verbs as in (2b) and (2d), and monologue prompts were varied between full-stop prompts (2a-b) and *because* prompts (2c-d):

- (2) a. Mary scolded_{IC} John. _____
 b. Mary saw_{nonIC} John. _____
 c. Mary scolded_{IC} John because _____
 d. Mary saw_{nonIC} John because _____

Of relevance here is what Kehler et al. found for the full-stop prompt condition (2a-b), specifically that IC contexts are more likely to give rise to Explanation continuations (60%) than a control set of non-IC verbs (24%). To put it in QUD terms, reading contexts like (2a) could be seen to bias the participant to answer the question *Why?*, whereas contexts like (2b) are biased toward answers to other possible questions (e.g., *What happened next?*).

We report here on the dialog component of the experiment, where we ask whether one will find evidence for a similar effect of context on expectancies in the QUD framework. We address this question by examining whether contexts bias participants toward asking particular questions in a dialog setting.

Participants were instructed to imagine a conversation with a friend and write a natural continuation that represented a question that they would be likely to pose (as in (3)).

- (3) Friend: Mary scolded_{IC} John.
 You: _____?

On analogy with the monologue condition, we expect that IC contexts will evoke more *Why?* questions than non-IC contexts, supporting the idea that addressees anticipate QUDs based on contextual cues.³

Methodology

Participants. 75 monolingual English-speaking undergraduates at UC San Diego (57 females; mean age=20) participated in the study for extra credit in Linguistics courses.

Materials and Procedure. A 2×3 design was employed that crossed verb type (IC verb or non-IC verb) with continuation type (full stop prompt, *because* prompt, or dialog prompt (3)). As noted above, the data collected for the full-stop prompt and *because* prompt conditions was previously reported in Kehler et al. (2008). The remaining data from the dialog condition now allows us to test whether QUDs are also sensitive to verb type.

For the experimental items, each context sentence mentioned two referents in a situation described with an IC verb or a non-IC verb. Forty IC verbs and forty non-IC verbs were taken from McKoon et al. (1993) with three replacements. The verbs *cheat*, *jeer*, and *dread* were felt to sound awkward in these sentence frames, and were replaced with *offend*, *mock*, and *fear*, respectively. The two referents consisted of one male and one female, with gender position balanced across items. To diversify the nature of the stimuli, twenty filler sentences used non-IC verbs and were followed by various interclausal connectives

³ Note that while we are interested in investigating the effect of IC verbs on the number of *Why?* questions, we are not addressing the question of what properties of these verbs (or of the eventualities they denote) are responsible for the effect. For discussion of this issue, see e.g. Hartshorne and Snedeker (2013), Bott & Solstad (2014), and Hartshorne et al. (2015), inter alia.

(monologue continuation) or a dialog response that contained the beginning of a question (dialog continuation), for a total of 100 items. Items were counterbalanced across six lists.

Discourse continuations were collected via a web-based interface that participants accessed from their own computer. Each item was presented on a page by itself with a text box in which participants were instructed to write their continuation. Items were presented to convey a dialog, with one line showing a friend's statement (the context sentence) followed by a prompt to elicit the continuation. The prompt started with "You:" and a question mark "?" after the text box signaled that a question was expected.

The entire experiment took roughly 45 minutes. Participants were instructed to imagine a natural discourse continuation for each prompt, writing the first one that came to mind and avoiding humor.

Annotation. Two trained judges (the second author and a Linguistics undergraduate) assessed the participants' continuations. To allow for a comparison with the monologue condition (see footnote 4), an annotation scheme was devised for categorizing QUDs in the dialog condition into groups based on the set of coherence relations listed in Table 1. For example, questions were annotated in terms of their possible answers such that a question like *Why?*, which asks for an answer in the form of an Explanation, would be categorized as an Explanation QUD (other questions that correspond to Explanation QUDs included *How come?*, *What for?*, or *For what reason?*). Examples of annotations are given in Table 2 and come directly from participants' story continuations for the experiment. The same annotation scheme was applied to both IC and non-IC contexts. Annotation disagreements were resolved by discussion.

Analysis. To measure the effect of verb type on the observed QUDs, we used mixed-effect logistic regressions (Jaeger, 2008). We modeled QUD as a binary choice (*Why?*-type question vs. non-*Why?*-type question) with a fixed-effect predictor for verb type (IC vs. non-IC). Verb type varied within participants and between items. For Experiment 1 and all subsequent logistic regressions in this paper, models included random intercepts and random slopes (Barr et al., 2013), with coefficient estimates and p-values reported for the centered fixed effect predictors (based on the Wald Z statistic; Agresti, 2002).

Results

After setting aside the responses in the dialog condition in which the participant misunderstood the task (interpreting the question prompt as a prompt to write a statement) or that were deemed to be ambiguous or non-information-seeking (rhetorical) (11.8% of the dialog responses), the remaining dataset for the mixed model contained 1788 dialog responses. All reported biases and figures use participant means. As predicted, IC contexts yielded more *Why?*-type questions (74.3%) than non-IC contexts (25.8%). In the mixed-effects model, verb type was a significant factor for modeling question types ($\beta = -2.589$, $p < 0.001$).⁴

Discussion

The goal of this experiment was to examine whether contextual cues affect expectancies for QUDs in a simple dialog condition, as measured by the types of questions that participants ask when presented with contexts involving different verb types. As predicted, verb type had a significant effect: Compared to non-IC contexts, IC contexts gave rise to significantly more *Why?*-type questions. Further, a correlation analysis revealed a pattern across the dialog and monologue conditions, whereby participants tend to ask the question in their dialog responses that they are likely to answer in their monologue continuations.

⁴ We also performed a correlation analysis between the distributions of coherence relations in the free-prompt monologue condition (2a-b) and of QUDs posed in the dialog condition. Because the mapping between coherence relations and QUDs is not one-to one (e.g., Violated-Expectations are somewhat common in monologue, but rarely posed as a question), we expect only a partial alignment between the percentages. If we include the 1514 full-stop monologue responses from Kehler et al. (2008) in our analysis, we can calculate the proportions for each participant or each item across verb types. A linear model of the proportion of *Why?*-type questions as a function of the proportion of Explanation relations across verb types, by participants and by items, show that the proportions (of *Why?*-type questions and Explanation coherence relations) are reliably correlated across the IC and non-IC conditions (Adjusted $R^2 = 0.37$, $F(1,148) = 88.39$, $p < 0.001$; Adjusted $R^2 = 0.44$, $F(1,77) = 63.23$, $p < 0.001$). This demonstrates that, at least to some extent, participants not only use contextual clues to anticipate ensuing QUDs, but also that they continue passages by answering those same QUDs.

Experiment 2: Effect of QUD Expectations on Pronoun

Interpretation during Self-Paced Reading

If comprehenders use contextual cues to form probabilistic expectations about the QUDs that the ensuing utterance is likely to answer, it stands to reason that these expectations would affect the processing of linguistic expressions that are dependent on those QUDs. Here we consider the problem of pronoun interpretation. The following simple equation captures the predictions of an expectation-driven QUD theory:

$$(4) \quad P(\text{pronoun} = \text{referent}) = \sum_{QUD \in QUDs} P(QUD) P(\text{pronoun} = \text{referent} | QUD)$$

Let us suppose, plausibly enough, that discourse contexts will give rise to probabilistic expectations about (i) what QUD the ensuing utterance will answer (as demonstrated by Experiment 1), represented as $P(QUD)$, and that (ii) expectations about pronominal referents are conditioned on what QUD the utterance that contains the pronoun is expected to answer, represented as $P(\text{pronoun} = \text{referent} | QUD)$. Equation (4) then predicts that contextual factors that affect expectations about QUDs (i.e., changing $P(QUD)$) will also affect expectations about pronoun interpretation (i.e., $P(\text{pronoun} = \text{referent})$). To take an example, say a discourse context effectively pushed the probability of the operative QUD being *Why?* to 1, for instance because the connective *because* occurred in the input. Then equation (4) predicts that the pronoun bias $P(\text{pronoun} = \text{referent})$ would be updated to be $P(\text{pronoun} = \text{referent} | QUD = \text{Why?})$. Being a law of conditional probability, of course, equation (4) is only interesting if pronoun interpretation is actually sensitive to QUDs, i.e., if $P(\text{pronoun} = \text{referent} | QUD)$ varies by QUD. If pronoun interpretation is primarily a ‘bottom up’ process that is sensitive only to explicit properties of the context (e.g., the grammatical roles occupied by constituents denoting potential referents, and so forth), expectations about QUDs should not matter.

We investigate this question by way of a self-paced reading study that uses transfer-of-possession

contexts. The study builds on previous passage completion experiments operating in the relational framework, which we now summarize, that have shown that pronoun interpretation biases in such contexts vary dramatically across coherence relations. First, a study reported on by H. Rohde et al. (2006; also included as Experiment 2 in Kehler et al. 2008) used transfer-of-possession context sentences followed by ambiguous pronoun prompts, as in (5):

(5) John handed a book to Bob. He _____

In the class of transfer verbs they used, the subject fills the Source thematic role and the object of the preposition fills the Goal role. Following Stevenson et al. (1994), Rohde et al. (2006) hypothesized that such verbs have a special property, specifically that the prominence of the two event participants varies across event structure: That is, the Source (which also acts as an Agent in being the participant carrying out the action) is central to the initial state and ongoing development of the event, whereas entity prominence shifts to the Goal – i.e., the recipient of the object of transfer – at the time the end state is reached. Rohde et al. thus predicted that coherence relations that focus on the end state of the previous eventuality (Occasion, Result) would display a strong Goal bias for pronoun interpretation, whereas relations that focus on other components of event structure (Explanation, Elaboration) would display a strong Source bias (see also Arnold, 2001). This is precisely what they found.

If pronoun interpretation is dependent on discourse coherence, then manipulations to the context that affect expectations about coherence should in turn affect expectations about pronoun interpretation. H. Rohde et al. (2007) report on another study that elicited passage completions with the same stimuli as H. Rohde et al. (2006), but added an additional manipulation that made the operative QUD explicit in the instructions. Specifically, in one version the participants were asked to have their completion answer the question *What happened next?*, whereas in the other they were to answer the question *Why?*. They hypothesized that the QUD *What happened next?* should lead to coherence relations that focus on the end state of the transfer event; the fact the Goal has greater prominence within the end state should in turn lead to an increased number of pronominal references to the Goal. Conversely, the QUD *Why?* should lead to coherence relations that focus on the initial state of the context event (since causes precede

effects); here the fact that the Source has greater prominence within the initial state should lead to more references to the Source. Hence the difference in QUDs was predicted to affect pronoun interpretation biases, even though the stimuli across conditions were identical. This was confirmed: There were significantly more Source interpretations in the *Why?* condition than the *What happened next?* condition.

Building on H. Rohde et al.'s findings, Experiment 2 tests whether the types of QUD biases observed in Experiment 1 are deployed rapidly in on-line comprehension, such that they would affect the processing of ambiguous pronouns. Like H. Rohde et al. (2007), we use the experimental instructions to guide comprehenders' expectations about the QUD that a follow-on sentence will answer with respect to the context sentence. The design is between-participants, so that half the participants received instructions explaining that they would read two-sentence passages in which the second sentence answers the question *Why?* relative to the first, which is predicted to lead to a Source bias for the pronoun. The other half received instructions stating that the second sentence would answer the question *What happened next?*, which is predicted to lead to a Goal bias for the pronoun. We adapted the stimuli from H. Rohde et al. (2007) to create a moving-window self-paced reading study with a 2×2 design that varied QUD expectations between participants (*Why?* vs. *What happened next?*) and coreference within participants (Source vs. Goal), as in (6)-(7). Underscores connect words that were presented together as a single region in the study.

- (6) [Source-referring pronoun] Jessica served chili to_Emily. She explained to Emily...
- a. [WHY] ... in_the_kitchen_that morning that_everyone needs_to try_chili_once.
 - b. [WHAT-NEXT] ... in_the_kitchen_that night that the_secret_to chili is_real_jalapenos.
- (7) [Goal-referring pronoun] Jessica served chili to_Emily. She explained to Jessica...
- a. [WHY] ... in_the_kitchen_that morning that_she can_only eat_soft_foods.
 - b. [WHAT-NEXT] ... in_the_kitchen_that night that the_chili was_a_bit too_spicy.

The critical region is the proper name in the second sentence in (6)-(7), as up to that point the pronoun

remains ambiguous. In order to disambiguate the pronoun, the comprehender must understand that the pronoun and proper name cannot be coreferential (e.g., the pronoun in (6) must refer to a referent other than the Goal *Emily*). This constraint is encapsulated in Condition C of Chomsky's (1981) Binding Theory: A non-pronominal referring expression must not be bound in its governing category (i.e., it must not co-refer with an element that structurally c-commands it, such as the subject pronoun in (6)-(7)). If comprehenders make use of Binding Theoretic constraints during processing, then we expect that biases comprehenders have regarding the interpretation of the sentence-initial pronoun should be detectable in their processing of the proper name, in that reading time should increase when the pronoun interpretation signaled by the proper name violates the favored interpretation. Because differences in processing difficulty in self-paced reading can show up a word or two downstream of the critical region (Mitchell, 1984), the immediately post-critical region (the spillover region) was always the same across conditions (*in the kitchen that* in (6)-(7)).

Our hypothesis predicts a QUD \times coreference interaction. In the *Why?* condition, we hypothesize that comprehenders will favor an interpretation of the pronoun as coreferential with the Source (because the initial state is associated with Source prominence), whereas in the *What happened next?* condition, we hypothesize that comprehenders will favor coreference with the Goal (because end states are associated with Goal prominence). Importantly, reading time is being measured in the critical/spillover regions, that is, before comprehenders encounter the remainder of the sentence and have sufficient information to reason about how the two sentences relate. Therefore, differences in reading time would be due to comprehenders' expectations regarding the relation that will hold between the two sentences. On the other hand, models of pronoun interpretation that rely primarily on grammatical (or thematic) role biases would predict uniform results across the two QUD conditions. For example, subject and parallel grammatical role preferences predict that the Source-referring pronoun in (6) would yield faster reading times in the critical/spillover regions than the Goal-referring pronoun in (7), whereas a thematic-role bias toward the Goal (Stevenson, et al., 1994; Arnold, 2001) predicts that the Goal-referring pronoun in (7) would yield faster reading times than the Source-referring pronoun in (6).

Methodology

Participants. Thirty-five monolingual English-speaking undergraduates from Northwestern University (19 female; mean age=20) participated for extra credit in Linguistics classes.

Materials. Each of the experimental items consisted of a first sentence describing a transfer-of-possession event and a second sentence containing a plausible continuation that either answered the question *Why?* or *What happened next?*. In the first sentence, the Source was mentioned in subject position and the Goal in a sentence-final prepositional phrase (both with proper names), and an inanimate Theme object followed the transfer verb. Multi-word themes (e.g. *a_gossipy_email*) were revealed as a single region, as was the sentence-final prepositional phrase containing the Goal reference (*to_Emily*). The second sentence consisted of a temporarily ambiguous sentence-initial pronoun followed by a verb that permits re-mention of either the Source or Goal referent as in (6) and (7). The re-mentioned Source/Goal appeared as a proper name either directly after the verb (e.g. *Mike handed a book to Josh. He heard Mike/Josh...*) or with one intervening word (e.g. *Angela forwarded a gossipy email to Kelly. She suspected that Angela/Kelly...*). The second sentence was presented word-by-word up to and including the proper name; subsequent words were presented as either single- or multi-word regions (maximum 5 words). The spillover region was always a multi-word region. The transfer verbs were the same as those used in H. Rohde et al. (2006, 2007). Items were counterbalanced across six lists.

The experiment consisted of six practice items, followed by 21 experimental items⁴ mixed with 40 fillers, pseudorandomized for each participant. Fillers were similar to the stimuli, consisting of two-sentence passages that provided coherent answers to the operative QUD. Because the QUD manipulation was between participants, the fillers differed between the *Why?* and *What happened next?* versions.

⁵ The use of 21 stimuli was a result of carrying over the verb set from Rohde et al. (2006), who had included a 3-way verb classification in their analysis.

Procedure. Participants read one of two versions of the experimental instructions, which explained that the items would consist of two-sentence passages in which the second sentence answered the question *Why?* or *What happened next?* relative to the first. Both versions included (non-transfer) examples.

Items were presented in a moving-window self-paced reading paradigm, using the Linger Software package (D. Rohde, 2003) on a Mac laptop with a 13” screen. Sentences initially appeared as a series of dashes (— — — —) obscuring the words, and participants pressed the space bar to reveal each region. Presentation was non-cumulative such that previous regions were replaced with dashes when the next region appeared. Participants pressed either a YES or NO button on the keyboard (‘f’ and ‘j’) to answer a comprehension question after every sentence, and they received automatic feedback whenever they answered incorrectly. They were instructed to read as quickly and carefully as possible, making sure they understood the sentences and slowing down if they answered multiple questions incorrectly. We recorded reading times for each region as well as the participant responses to the comprehension question. Two breaks were automatically offered during the experiment.

Analysis. For comprehension question accuracy, we used mixed-effect logistic regressions. Accuracy was modeled as a binary outcome (correct vs. incorrect) with centered fixed predictors for QUD type (*Why?* vs. *What next?*) and coreference (Source-referring pronoun vs. Goal-referring pronoun). For the reading time (RT) data, we analyze residual RTs at the critical region and the spillover region. Residual RTs adjust for overall differences in participants’ reading rates as well as differences in readers’ sensitivity to word length. They were calculated as the difference between the actual reading time on a word and the reading time predicted by a regression equation (computed separately for each participant, using all experimental and filler items) relating word length to reading time (Trueswell, Tanenhaus, & Garnsey, 1994). RTs were modeled using mixed-effect linear regressions, with likelihood ratio tests to compare mixed-effects models differing only in the presence or absence of the fixed factors and their interaction. Fixed factors were centered, and models included random intercepts and slopes.

Results

After excluding one participant whose comprehension-question accuracy on experimental items was not significantly better than chance, the mean accuracy was 95.1% for fillers and 87.0% for experimental items, indicating that participants paid attention to the task.

Comprehension question accuracy. Table 3 shows the mean accuracy on comprehension questions for each condition. In the tables and figure in this section, the conditions are annotated with the referent of the pronoun after disambiguation (not the referent of the re-mentioned proper name at the critical region) and the QUD expectation established by the instructions (e.g., **Pronoun=Source**; **Why** indicates a Source-referring pronoun in the *Why?* condition: *Jessica served chili to Emily. She_{SOURCE} explained to Emily...*). Neither of the main effects nor the QUD \times coreference interaction was a significant factor for modeling question accuracy (QUD: $\beta=-0.549$, $p=0.172$; coreference: $\beta=-0.150$, $p=0.72$; QUD \times coreference interaction: $\beta=0.652$, $p=0.41$).

Reading time results. Table 4 shows the raw reading times by condition for the critical and spillover regions. Figure 1 shows residual reading times by condition starting in the first sentence.

For the analysis, we model residual RTs, regardless of comprehension-question accuracy. Residual RTs more than three standard deviations away from the mean, per region and per condition (<1% of the data), were replaced with the maximum RT for that region and condition.

At the disambiguating proper name (*Emily/Jessica*), RTs showed no main effects of QUD expectations ($\beta=-14.10$, $p=0.62$) or coreference ($\beta=13.04$, $p=0.52$), but there was a significant QUD \times coreference interaction in the predicted direction ($\beta=102.02$, $p<0.05$): In the *Why?* condition, names that signaled Source-referring pronouns were read more quickly than names that signaled Goal-referring pronouns, whereas this was reversed in the *What happened next?* condition. To test for a full crossover interaction, we conducted pairwise analyses. In the Source-reference condition, the effect of QUD was not significant ($\beta=37.99$, $p=0.25$). In the Goal-reference condition, the *What happened next?* condition

yielded reliably faster RTs than the *Why?* condition ($\beta=-65.81$, $p<0.001$).

At the spillover region (*in the kitchen that*), the only significant result was a main effect of coreference, whereby names that signaled Source-referring pronouns were read more quickly than names that signaled Goal-referring pronouns ($\beta=-94.779$, $p<0.05$). There was no effect of QUD ($\beta=1.257$, $p=0.24$) nor a QUD \times coreference interaction ($\beta=58.651$, $p=0.18$). Analyses of the raw RTs and analyses of the residual RTs with incorrectly answered items excluded were qualitatively the same. No pre-critical regions yielded significant effects.

Discussion

Experiment 2 was designed to test the hypothesis that expectations about the operative QUD can influence on-line pronoun interpretation. This hypothesis was confirmed by the significant interaction between QUD type and coreference on reading times at the critical region.

However, the nature of the interaction reveals that it was only the passages with Goal-referring pronouns that were processed according to our predictions—proper names that signaled Goal-referring pronouns were read faster in the *What happened next?* condition than the *Why?* condition. On the other hand, the results for passages with Source-referring pronouns, while numerically in the predicted direction, failed to reach significance. One possible explanation for this result is the known asymmetry between subject and nonsubject referents with respect to the speaker's choice of referring expression. Specifically, recent work has shown that speakers pronominalize references to the subject referent considerably more often than other grammatical roles, and further that, rather counter-intuitively, this bias appears to be insensitive to the semantic and pragmatic factors that comprehenders bring to bear during interpretation (H. Rohde, 2008; Fukumura & Gompel, 2010; H. Rohde & Kehler, 2014; cf. Arnold 2001). Indeed, the failure to pronominalize a subject referent mention can lead to a so-called repeated name penalty whether or not the preferred referent for a pronoun would have been the subject (Gordon, Grosz, & Gilliom, 1993; Gordon & Searce, 1995; Almor, 1999). Fukumura and van Gompel (2010) explicitly

link the repeated-name penalty with the strong production bias to pronominalize a re-mention of the previous subject: The processor is so surprised to see a name instead of a pronoun for a referent in subject position that it causes a reading time delay. If this is correct, it stands to reason that the processor would therefore *not* be surprised to see a pronoun used to refer to a subject referent even if anticipatory semantic and pragmatic biases were pointing it away from that referent, hence leading to a lack of processing delay for bias-incongruent subject references. Because this reasoning only applies to subject referents, the use of a name to refer to nonsubject is not similarly constrained. Referring to the nonsubject Goal referent (in our case with a pronoun) is therefore predicted to be less expected in the *Why?* condition than the *What happened next?* condition due to the differing interpretation biases, which was confirmed by our study.⁶

Experiments 3a and 3b: Effect of Pronoun Prompt on QUD Expectations

The previous study demonstrated that expectations concerning what QUD a subsequent utterance will answer affect referential biases. The results fit well with a model in which QUD expectations influence the incremental interpretation of linguistic expressions. A further prediction of a fully incremental model is that the interpretation of certain linguistic expressions can in turn influence QUD expectations. That is, if QUD expectations influence referential processing (Experiment 2), we now ask whether referential processing can influence QUD expectations (Experiments 3a/b).

On analogy with equation (4), the predictions of an expectation-driven QUD theory regarding the effect of reference on QUDs is captured by equation (8):

⁶ An anonymous reviewer points out that stimuli such as (7), in which the pronoun is Goal-referring, violate Rule 2 of Centering theory (Grosz, Joshi, and Weinstein, 1995), since a repeated name is used to refer to the previous subject (the presumptive backward-looking center, or C_b), whereas a pronoun is used to refer to a non- C_b . This is not true of stimuli like (6), in which the pronoun is Source-referential. This may explain why a slowdown was found in the spillover region of the Goal-referential stimuli as compared to the Source-referential ones. It does not explain our interactive effects, however, since the Goal-referential stimuli in both instruction conditions were the same in this respect. Indeed, as indicated above, the Goal-referential stimuli in the *What happened next?* condition were read the fastest of all four conditions.

$$(8) \quad P(QUD) = \sum_{referent \in referents} P(referent) P(QUD | referent)$$

To see how this might work, first consider the predicted effect of encountering an unambiguous pronoun after an object-biased IC context sentence as in (9a-c):

- (9) a. Mary scolded John. _____
 b. Mary scolded John. He _____
 c. Mary scolded John. She _____

After processing the context sentence, a comprehender will have a particular distribution $P(QUD)$ representing his expectations about ensuing QUDs. We know from Experiment 1 that this distribution will vary across different contexts; for instance IC contexts such as (9a) are biased in favor of the question *Why?* However, per equation (8), we would expect that encountering a subsequent unambiguous referring expression would alter these expectations. For instance, when the comprehender then encounters a sentence-initial reference to the subject as in (9c), the probability of $P(referent=subject)$ will change to 1. Per equation (8), this increase will cascade to lower the probability for QUDs for which $P(QUD | referent=subject)$ is low in favor of QUDs for which this probability is higher; for example, since *scold* in (9) is an object-biased IC verb, we would expect a subject reference to lead to a drop in the likelihood of a *Why?* continuation. By the same reasoning, an unambiguous reference to the object as in (9b) should have the effect of raising the likelihood of a *Why?* continuation for such verbs.

We can now ask, on the expectation-driven QUD model, what would happen if an ambiguous pronoun were encountered rather than an unambiguous one, as in (10)?

- (10) Bob scolded John. He _____

One might naively predict that the pronoun will have no effect on QUD expectations, since the pronoun could refer to either event participant. However, recall from the discussion of Experiment 2 that production studies have found that references to a previous subject are pronominalized more often than references to entities that occupy other grammatical roles. This means that the speaker's choice to use a

pronoun is expected to contribute a bias such that the probability of $P(\text{referent}=\text{subject})$ will rise. (This effect has been confirmed by passage completion studies that have shown that ambiguous pronoun prompts yield more continuations about the subject than full-stop prompts; Stevenson et al., 1994; H. Rohde, 2008; H. Rohde & Kehler, 2014). Equation (8) then predicts that this increase will cascade to raise the probability for QUDs for which $P(\text{QUD} \mid \text{subject})$ is high, and correspondingly reduce the probability for QUDs for which $P(\text{QUD} \mid \text{subject})$ is low. The prediction is completely analogous to the situation for the unambiguous pronoun in (9c), except that an ambiguous pronoun will cause $P(\text{referent}=\text{subject})$ to merely rise rather than to become 1.

As such, the expectation-driven QUD model predicts a greater percentage of continuations that employ subject-biased QUDs in a pronoun-prompt condition as compared to a full-stop prompt condition. This reasoning assumes, of course, that QUD biases are updated incrementally. If QUD expectations are independent of reference, then we should be just as likely to see answers to the question *Why?* in (10) as we would if the pronoun did not appear in the prompt. A result revealing the predicted effect would therefore establish that the relationship between reference and QUDs is bidirectional: Whereas Experiment 2 established that QUDs influence pronoun interpretation expectations, a positive outcome here would show that pronoun interpretation biases external to QUDs influence QUD expectations.

To test this prediction, we had participants complete passages in pronoun-prompt and full-stop conditions in two studies. The first (Experiment 3a) employs an IC verb-type manipulation. To ensure the effects found are not limited to IC manipulations, the second (Experiment 3b) extends the result by examining the effect of a grammatical manipulation (aspect) in transfer-of-possession contexts.

Experiment 3a: Effect of Pronouns in Implicit Causality Contexts

The first experiment examines the hypothesis using IC contexts.⁷ Methodologically, asking

⁷ Data regarding reference production and interpretation from this experiment was previously reported in H. Rohde and Kehler (2014). Here we analyze the data with respect to our hypotheses regarding the distribution of coherence relations in participant continuations, which were not addressed in that work.

participants to provide questions in a dialog condition as we did for Experiment 1 would be awkward, since in the pronoun prompt condition they would have to posit a question that intervenes between the context and the pronoun-initial clause, a rather unnatural task. As such, we instead ask participants to complete the passage, and analyze the data in terms of coherence relations. Because an effect on the distribution of coherence relations can be mapped to an effect on the distribution of their corresponding QUDs (e.g., a shift in the distribution of Explanation relations entails a corresponding shift in the distribution of *Why?* questions), the coherence analysis can be used to evaluate the effect of a pronoun prompt on expectations concerning ensuing QUDs. In what follows, we will therefore continue to speak in terms of QUDs when describing the predictions of the analysis, but will speak in terms of coherence relations when describing the annotation and data analysis.

The stimuli employed the same contexts used in Experiment 1, with changes made so that they contain two same-gender referents and with the prompt type varied between full-stop and ambiguous pronoun versions, as in (11)-(13).

- | | | |
|------|------------------------------------|----------|
| (11) | John infuriated Bob. (He) _____. | [IC-1] |
| (12) | John scolded Bob. (He) _____. | [IC-2] |
| (13) | John chatted with Bob. (He) _____. | [non-IC] |

To test our hypothesis, we compare the distribution of coherence relations in the pronoun prompt and full-stop prompt conditions. The prediction is that the ambiguous pronoun prompt will yield more subject-biased relations than object-biased ones, as compared to the proportions in the full-stop condition.⁸ Depending on the verb class (IC-1, IC-2, non-IC), different relations are significantly affected by

⁸ To be clear, this does not mean we expect that the overall pronoun interpretation bias will be toward the subject, and hence yield an overall bias toward subject-biased coherence relations, particularly for IC-2 contexts. The prediction is that including a pronoun in the prompt will lead to more first-mentions of the subject when compared to the full-stop condition, and that this will lead to comparatively more subject-biased relations.

coreference, so the analysis must be conducted separately by verb class.

Methodology

The task and procedure follow that of Experiment 1.

Participants. Twenty-eight monolingual English-speaking undergraduates from UC San Diego (17 females; mean age=20.5) participated either for extra credit in Linguistics courses or for the chance to win a gift certificate.

Materials. The same verbs were used as in Experiment 1. Context sentences for the experimental items mentioned two referents, either both male or both female, with gender balanced across items. Participants saw each verb only once, in one of the two prompt conditions. The experiment consisted of 100 items: 80 experimental items (40 IC, 40 non-IC) intermixed with 20 non-IC fillers. The additional fillers contained non-IC verbs followed by intersentential connectives, full-stop, or pronoun prompts.

Evaluation and Analysis. Two judges (the second author and a Linguistics undergraduate) annotated the participants' continuations for reference and coherence relations. For the hypothesized effect of prompt type on the distribution of coherence relations, the three verb classes are analyzed separately. We use mixed-effect logistic regressions, modeling the coherence outcome as a binary choice (subject-biased relation vs. non-subject-biased relation).

The identity of subject-biased and non-subject-biased relations was determined independently using the data reported in Kehler et al.'s (2008) Experiment 3. As explained above, equation (8) predicts that the introduction of a subject bias will shift the distribution for those relations for which $P(\text{QUD} \mid \text{subject})$ is significantly different from $P(\text{QUD} \mid \text{object})$ for their corresponding QUDs; those for which $P(\text{QUD} \mid \text{subject})$ is significantly higher are hence included in the subject-biased category, and those for which $P(\text{QUD} \mid \text{object})$ is higher are included in the object-biased category. The analysis of the Kehler et al. data revealed that the relations whose probabilities were significantly conditioned by coreference for IC-1 verbs are Explanation ($\beta=3.066$, $p<0.001$; subject-biased), Occasion ($\beta=-3.772$, $p<0.005$; object-

biased), and Result ($\beta = -3.885$, $p < 0.001$; object-biased). Analysis of other relations yielded non-significant effects (Elaboration: $\beta = 0.935$, $p = 0.09$; Parallel: $\beta = 10.500$, $p = 0.28$; Violated Expectation: $\beta = 14.046$, $p = 0.33$). The IC-1 analysis will therefore test for a main effect of prompt type on the proportion of subject-biased versus object-biased relations (Explanation vs. Occasion and Result). For the analysis of IC-2 verbs, we include Elaborations ($\beta = 1.939$, $p < 0.001$; subject-biased) and contrast them with Explanations ($\beta = -0.980$, $p < 0.05$; object-biased) and Results ($\beta = -2.891$, $p < 0.001$; object-biased); we again exclude the relations whose use does not depend significantly on reference (Occasion: $\beta = -22.17$, $p = 0.60$; Parallel: $\beta = -15.26$, $p = 0.996$; Violated Expectation: $\beta = 0.394$, $p = 0.52$). For the analysis of non-IC verbs, we include Elaborations ($\beta = 0.711$, $p < 0.05$; subject-biased) and Explanations ($\beta = 1.009$, $p < 0.005$; subject-biased) and contrast them with Results ($\beta = -1.142$, $p < 0.005$; object-biased). We again exclude relations with non-significant effects of reference (Occasion: $\beta = -0.556$, $p = 0.28$; Parallel: $\beta = -2.267$, $p = 0.44$; Violated Expectation: $\beta = -0.427$, $p = 0.21$).

Results

The analysis reported here represents a conservative analysis in which continuations were excluded if at least one coder assessed the reference or coherence relation to be ambiguous (15.7% of continuations). For continuations elicited in the full-stop condition, the analysis also excludes continuations in which the participant failed to mention one of the referents in the context sentence or used a conjoined noun phrase (*John and Bob*), a plural pronoun (*they*), or a possessive (*his*) (16.6% of continuations). The remaining dataset contained 1516 continuations.

To evaluate whether the pronoun prompt condition induces a preference for subject-biased coherence relations,⁹ the analysis for each verb class is restricted to continuations whose coherence

⁹ As reported in Rohde and Kehler (2014), the analysis of coreference confirms the predicted higher rates of (i) pronominalizations of re-mentions of subject referents as compared to nonsubject referents in the no-pronoun condition, and (ii) continuations about the previous subject in the pronoun prompt condition as compared to the no-pronoun condition.

relations are sufficiently polarized in their coreference bias (see *Evaluation and Analysis* above). For both IC verb classes, prompt type exhibits the predicted effect on the proportion of subject-biased coherence relations (IC-1: $\beta=6.068$, $p<0.001$; IC-2: $\beta=1.598$, $p<0.001$). The results for non-IC verbs however, while numerically in the predicted direction, failed to reach significance ($\beta=0.531$, $p=0.157$). The effects of pronoun prompts on the percentage of subject-biased relations compared to full-stop prompts is shown for each verb class in Figure 2.

Discussion

In the case of IC-1 and IC-2 verbs, the differing coherence distributions in the full-stop and pronoun prompt conditions show that biases associated with pronouns are factored into participants' expectations about what QUD the ensuing clause will answer. As predicted, the mere presence of an ambiguous pronoun shifts these expectations toward subject-biased QUDs. For the case of non-IC verbs, the difference between coherence distributions, while numerically in the right direction, failed to reach significance. An analysis of the data, however, suggests that this result may not be surprising. For these verbs, Occasion is the third most common coherence relation, but Occasion continuations were not included in the analysis since for this relation the (rather heterogeneous) collection of non-IC verbs used did not display a strong dependence on referent. This in turn caused the analysis to lose data. A second examination is therefore required to ensure that the effect identified here generalizes beyond IC verbs; one that utilizes a more homogenous set of contexts. The design of Experiment 3b satisfies this desideratum.

Experiment 3b: Effect of Pronouns in Transfer-of-Possession Contexts

As a second test, we now examine the hypothesis using a different context type, specifically one that contains Source-Goal transfer-of-possession verbs. Unlike Experiment 3a, which manipulated context by varying verb type, Experiment 3b manipulates a grammatical property, specifically aspect. Our design is motivated by the results of Kehler et al. (2008), Experiment 2, which showed that

manipulating the verbal aspect of a context sentence expressing a transfer-of-possession event shifts the distribution of coherence relations. First, aspect influences attention with respect to event structure: The perfective (*handed*) is used to describe an event as completed, whereas the imperfective (*was handing*) is used to describe an event as ongoing. Rohde et al. found that perfective verbs were followed more frequently by Occasion relations than their imperfective counterparts, since Occasion is sensitive to the end state of the previous event. Imperfective verbs were followed more frequently by Elaborations and Explanations, on the other hand, since they are sensitive to an event's initiating conditions, and are hence less dependent on the event being described as completed.

For this study a 2×2 design was employed, crossing verb aspect (perfective or imperfective) and prompt type (pronoun prompt or full-stop prompt):

(14) John gave a book to Bill. _____

(15) John gave a book to Bill. He _____

(16) John was giving a book to Bill. _____

(17) John was giving a book to Bill. He _____

As in Experiment 3a, judges annotated the passages for reference and coherence relations. We test for a main effect of prompt type on the proportions of coherence relations biased to the subject and nonsubject, whereby a pronoun prompt is predicted to yield a greater number of subject-biased relations than a full-stop prompt.

Unlike Experiment 3a, the status of the coherence relations as biased to the subject or nonsubject is the same across both verb aspects, so the coherence analysis can be conducted on the entire dataset at once. We use the data from Kehler et al. (2008), Experiment 2, to determine which relations are associated with distributions that significantly differ when conditioned on a subject or nonsubject referent. Logistic regressions showed a main effect of referent on the presence/absence of five relations: Elaboration ($\beta=8.337$, $p<0.001$; subject-biased), Explanation ($\beta=2.605$, $p<0.001$; subject-biased), Violated Expectation ($\beta=1.842$, $p<0.001$; subject-biased), Occasion ($\beta=-3.921$, $p<0.001$; object-biased),

and Result ($\beta = -5.701$, $p < 0.001$; object-biased). The effect for Parallel relations, which were infrequent, was not significant ($\beta = -6.035$, $p = 0.49$). The analysis here will therefore test for a main effect of prompt type on the proportion of subject-biased versus nonsubject-biased relations (Elaboration, Explanation, and Violated Expectation vs. Occasion and Result).

We also expect to replicate the main effect of aspect on the coherence proportions reported by Kehler et al. (2008). As aspect and prompt type are independent factors, there is no reason to expect an interaction between them.

Methodology

The task and procedure follow that of Experiment 1.

Participants. Fifty-two monolingual English-speaking undergraduates from UC San Diego (38 females; mean age=20.7) participated for extra credit in Linguistics courses.

Materials and Procedures. The contexts from Kehler et al. (2008), Experiment 2, were used, all containing a transfer-of-possession context with two possible referents of the same gender. The experiment consisted of 63 items: 21 experimental items interleaved with 21 stimuli for an unrelated experiment and 21 additional fillers. The stimuli for the interleaved experiment consisted of a matrix clause followed by the beginning of a relative clause with an ambiguous relative pronoun as the story continuation prompt. The additional fillers consisted of sentences with non-transfer verbs and a variety of prompts (pronouns, proper names, relative pronouns, adverbs). Items were counterbalanced across four lists.

Evaluation and Analysis. The coding procedure and statistical analysis followed that used in Experiment 3a. Unlike Experiment 3a, however, the status of the coherence relations as biased to the subject or nonsubject is the same across both verb aspects, so as indicated earlier, the coherence analysis can be conducted on the entire dataset at once.

Results

The analysis represents a conservative analysis in which continuations were excluded if at least one coder assessed the coherence or coreference relation to be ambiguous (10.8% of continuations). For continuations elicited in the full-stop condition, the analysis also excludes continuations in which the participant failed to mention one of the referents in the context sentence or used a conjoined noun phrase (*John and Bob*), a plural pronoun (*they*), or a possessive (*his*) (8.5% of continuations). The remaining dataset contained 881 continuations.

We first confirm that our data replicates the difference in rates of pronominalization for subject and nonsubject referents in the full-stop condition. As predicted, there is a main effect of grammatical role of the referent ($\beta=11.473$, $p<0.001$): Participants produce more pronouns when re-mentioning a subject referent (78.6%) than an object referent (10.7%). There is also a main effect of aspect due to the increased proportion of subject references (and hence pronominalized subject references) in the imperfective condition ($\beta=7.059$, $p<0.05$) but no grammatical role \times aspect interaction ($\beta=-2.540$, $p=0.65$). The full dataset also confirms that the proportion of continuations about the subject differs by prompt type (in a model crossing prompt type and verb class, main effects of prompt type: $\beta=2.750$, $p<0.001$; aspect: $\beta=1.360$, $p<0.001$; no interaction: $\beta=0.362$, $p=0.44$): Participants wrote more continuations about the subject in the condition with the pronoun prompt (62.1%) than the full-stop prompt (22.6%).

To evaluate the prediction that a pronoun prompt induces a preference for subject-biased coherence relations compared to the full-stop case, we model the coherence outcome as a binary choice (subject-biased relation vs. non-subject-biased relation). A main effect of prompt type ($\beta=1.462$, $p<0.001$) confirms our predictions that pronoun prompts favor subject-biased relations (71.6%) to a greater extent than than full-stop prompts (53.5%). There is also a main effect of aspect ($\beta=-1.277$, $p<0.001$), which replicates previous work showing that imperfective aspect induces more continuations in which the

operative coherence relation is an Elaboration, Explanation, or Violated Expectation (71.9%) than perfective aspect (53.4%). There is no prompt \times aspect interaction ($\beta=-0.565$, $p=0.18$). The results are shown in Figure 3.

Discussion

The results of Experiment 3b again confirmed the hypothesis concerning the effect of referential form on QUD expectations. Specifically, the predictions of equation (8) were borne out: Encountering a pronoun in the continuation clause introduced a bias toward a subject mention, which then led to more subject-biased QUDs as compared to the control condition in which no pronoun was given. The effect of prompt type on coherence therefore shows that participants appear to incorporate these pronominalization biases as they generate expectations regarding upcoming QUDs in transfer-of-possession contexts, again demonstrating that the dependency between QUDs and coreference is bidirectional.

General Discussion

The hypothesis that discourses are structured by implicit QUDs makes a series of interrelated predictions when situated within a dynamic, incremental model of sentence and discourse processing. First, it predicts that comprehenders will have anticipatory expectations about the QUDs that a subsequent utterance will answer based on properties of the context. This was confirmed by Experiment 1, a dialog completion study that showed that the verb type employed in a context sentence had a significant impact on the questions that participants followed up with. Furthermore, the questions displayed substantial alignment with the implicit questions that their completions answer in a monologue version of the task. Whereas this pattern is expected on an expectation-driven QUD model of discourse interpretation, it receives no natural explanation on a purely reactive model of discourse processing, in which comprehenders interpret an utterance before reasoning about the QUD to which it provides an answer.

Second, the hypothesis predicts that expectations about QUDs will affect interpretation biases associated with discourse-dependent linguistic phenomena. This was confirmed by Experiment 2, a self-

paced reading study, that showed that Goal-referential pronouns in transfer-of-possession contexts create greater processing difficulty when the operative QUD is *Why?* than *What happened next?*. This behavior is predicted by a model in which linguistic forms in an utterance are interpreted based in part on expectations about the QUDs that the utterance will answer. This pattern is difficult to explain, on the other hand, on a model in which pronoun interpretation is governed by factors independent of expectations concerning the operative QUD.

Third, the hypothesis predicts that independent biases associated with linguistic material will in turn affect expectations about the operative QUD. This was confirmed by Experiments 3a/b, which showed that the appearance of an ambiguous pronoun affects the distribution of coherence relations in, and hence of QUDs answered by, the continuations participants provided. The effect is predicted by a model in which the interpretation of pronouns and the identification of implicit QUDs are jointly inferred, such that the independent biases contributed by a pronoun cascade to influence biases toward different QUDs. This pattern is difficult to explain, on the other hand, if comprehenders wait until they resolve ambiguities on independent grounds before initiating reasoning about what QUDs utterances answer.

The results of these experiments therefore accord with the predictions of an expectation-driven QUD model of sentence and discourse processing. A remaining question, one not answered here, pertains to whether QUD theories of coherence are to be favored over other theories, e.g. a coherence-relation analysis that posits the existence of content-bearing relations between clauses. Whereas it seems clear that there is some overlap between the two analyses of discourse coherence, there are important differences as well. For instance, Roberts (2012/1996) notes briefly that certain coherence relations can be characterized in terms of implicit questions; for instance, we have already noted that example (1) can be straightforwardly treated as having the implicit question *Why?* intervening between the clauses on the QUD analysis or as an Explanation relation on the relational approach. Note that the constraints on recognizing the coherence of this discourse are similar on the two analyses: Recovering an implicit Explanation relation or an implicit QUD *Why?* both require that the hearer use linguistic clues and world knowledge to infer that the second sentence of (1) describes the *reason* for the event described in the first sentence. Another relation with a direct correspondence is Parallel (see Table 1), which can be treated as

relating a set of utterances that each provide a partial answer to a common QUD (e.g., *Andy read an article and Hannah ran some statistics*, in answer to the question *What did the co-authors do this morning?*). As we have noted, other corresponding relationships between coherence relations and questions exist: *How come?* and *What for?* also correspond to Explanation, *What happened next?* corresponds to Occasion, *Where/when/how?* correspond to Elaboration, and so forth. On the other hand, it is less common to find questions that correspond directly to relations such as Violated-Expectation in discourse, for example, insofar as interlocutors are less likely to inquire about specific events that contrast with expectation. Further work will be necessary to tease apart the differing predictions of the two approaches and determine which best represents the process by which the coherence of discourse is established.

An early proponent of the QUD analysis might have been the philosopher R.G. Collingwood, who once stated that “Every statement that anybody ever makes is made in answer to a question” (Collingwood, 1940; von Fintel, 2004). If QUDs are the central structuring devices that give discourses their coherence, then it stands to reason that theories of discourse-dependent linguistic expressions need to account for them. Indeed, QUDs have been utilized in analyses of a variety of linguistic phenomena including (but not limited to) accent placement (Roberts, 1996; Büring, 2003), domain restriction (Fintel, 2004), VP ellipsis (Kehler & Büring, 2007; Miller & Pullum 2014; Miller & Hemforth 2015; Kehler 2015), focus sensitivity (Beaver & Clark, 2008), and the meaning of discourse connectives (Toosarvandani, 2014). Among the contributions of this paper is the demonstration that pronoun interpretation is likewise sensitive to QUDs, and hence that incorporating QUDs into theories of pronoun processing and interpretation is necessary if those theories are to be empirically adequate. This observation carries important theoretical implications for pronoun research, in that factors that are known independently to influence QUD interpretation may in turn have repercussions for pronoun interpretation, even though the causal relationship between such factors and pronoun interpretation is indirect (for analyses of the relationship between accent placement and pronoun interpretation, for instance, see Kehler (2005) and Cummins & Rohde (2015)). More generally, the results presented here motivate extensions to dynamic theories of QUD-structured discourse interpretation to not only represent the space of possible

implicit questions, but also probability distributions over them (and for that matter, over their possible answers), and ultimately over the space of strategies of inquiry within which those questions are situated.

Acknowledgments

This research was supported by an Andrew W. Mellon postdoctoral fellowship to Hannah Rohde. We thank research assistants Erica Gold, Brittany Young, and Meredith Larson for their work in annotating and conducting the studies.

References

- Anderbois, S. (2011). *Issues and Alternatives*, PhD Thesis, University of California, Santa Cruz.
- Agresti, A. (2002). *Categorical data analysis* (second ed.). Wiley.
- Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106(4), 748-765.
- Altmann, G. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191-238.
- Arnold, J. E. (2001). The effects of thematic roles on pronoun use and frequency of reference. *Discourse Processes*, 31, 137-162.
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.
- Barr, D. J., Levy, R., Scheepers, C. and Tily H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3): 255–278.
- Beaver, D. I., & Clark, B. Z. (2008). *Sense and sensitivity: How focus determines meaning*. West Sussex, UK: Wiley-Blackwell.
- Bott, O., & Solstad, T. (2014). From verbs to discourse: A novel account of implicit causality. B. Hemforth, B. Mertins, & C. Fabricius-Hansen (Eds.), *Psycholinguistic approaches to meaning and understanding across languages* (p. 213-251). Springer.
- Büring, D. (2003). On d-trees, beans, and b-accent. *Linguistics and Philosophy*, 26, 511-545.
- Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns.

- Journal of Verbal Learning and Verbal Behaviour*, 16, 601–609.
- Carlson, L. (1983). *Dialogue games: An approach to discourse analysis*. Dordrecht: Reidel.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Collingwood, R. G. (1940). *An essay on metaphysics*. Oxford: Clarendon Press.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: the use of syntax by the psychological syntax processor. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing* (p. 320-358). Cambridge UK: Cambridge University Press.
- Cummins, C. & Rohde, H. (2015). Evoking context with contrastive stress: Effects on pragmatic enrichment. *Frontiers in Psychology, Special Issue on Context in Communication: A Cognitive View*, 6(1779), 1-11.
- Fintel, K. von. (2004). A minimal theory of adverbial quantification. In B. Partee & H. Kamp (Eds.), *Context dependence in the analysis of linguistic meaning* (p. 137-175). Elsevier.
- Fukumura, K., & Gompel, P. G. van. (2010). Choosing anaphoric expressions: Do people take into account likelihood of reference? *Journal of Memory and Language*, 62, 52-66.
- Garvey, C., Caramazza, A., & Yates, J. (1976). Factors underlying assignment of pronoun antecedents. *Cognition*, 3, 227-243.
- Ginzburg, J. (1996). Dynamics and the semantics of dialog. In J. Seligman (Ed.), *Language, logic and computation, volume 1*. Stanford: CSLI.
- Ginzburg, J., & Sag, I. (2000). *Interrogative investigations*. Stanford: CSLI Publications.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17, 311-347.
- Gordon, P. C., & Searce, K. (1995). Pronominalization and discourse coherence, discourse structure and pronoun interpretation. *Memory and Cognition*, 23, 313-323.
- Hartshorne, Joshua K. & Jesse Snedeker. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, 28(10), 1474-1508.
- Hartshorne, Joshua K., Timothy J. O'Donnell, & Joshua B. Tenenbaum (2015). The causes and

- consequences explicit in verbs. *Language, Cognition, and Neuroscience* 30(6), 716-734.
- Hobbs, J. R. (1990). *Literature and cognition, lecture notes 21*. Stanford, CA: CSLI.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* (Special issue on Emerging Data Analysis), 59, 434-446.
- Kehler A. (2005). Coherence-driven constraints on the placement of accent. In *Proceedings of Semantics and Linguistic Theory 15*. Los Angeles, CA.
- Kehler A. (2015). On QUD-based licensing of strict and sloppy ambiguities. In *Proceedings of Semantics and Linguistic Theory 25*. Stanford, CA.
- Kehler, A., & Büring, D. (2007). Be bound or be disjoint! In *Proceedings of the 38th Meeting of the North East Linguistics Society (NELS-38)*. Ottawa, Canada.
- Kehler, A., Kertz, L., Rohde, H., & Elman, J. (2008). Coherence and coreference revisited. *Journal of Semantics*, 25, 1-44.
- Koornneef, A. W., & van Berkum, J. J. A. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye-tracking. *Journal of Memory and Language*, 54, 445-465.
- Klein W., & von Stutterheim, C. (1989). Referential Movement in Descriptive and Narrative Discourse. In R. Dietrich & C. F. Graumann (eds), *Language Processing in Social Contexts*. North Holland: Elsevier Science Publishers B.V.
- Kuppevelt, J. van. (1995). Discourse structure, topicality, and questioning. *Journal of Linguistics*, 31, 109-147.
- Larsson, S. (1998). Questions under discussion and dialogue moves. In *Proceedings of TWLT13/ TWENDIAL '98: Formal semantics and pragmatics of dialogue*.
- McKoon, G., Greene, S., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology*, 19, 1040-1052.
- Miller, P. & Pullum, G.K. (2014). Exophoric VP Ellipsis. In Philip Hofmeister & Elizabeth Norcliffe

- (eds.), *The Core and the Periphery: Data-Driven Perspectives on Syntax Inspired by Ivan A. Sag*. 5–32. Stanford, CA.
- Miller, P. & Hemforth, B. (2015). VP Ellipsis Beyond Syntactic Identity: The Case of Nominalized Antecedents. Submitted for publication.
- Mitchell, D. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (p. 6989). Hillsdale, NJ: Earlbaum.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics, 49: Papers in Semantics*.
- Rohde, D. (2003). *Linger: a flexible platform for language processing experiments*, version 2.94. <http://tedlab.mit.edu/~dr/Linger/>.
- Rohde, H. (2008). *Coherence-driven effects in sentence and discourse processing*. Unpublished doctoral dissertation, UC San Diego.
- Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, to appear.
- Rohde, H., Kehler, A., & Elman, J. (2006). Event structure and discourse coherence biases in pronoun interpretation. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Vancouver, July 26-29, 2006.
- Rohde, H., Kehler, A., & Elman, J. (2007). Pronoun interpretation as a side effect of discourse coherence. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society*. Nashville, TN, August 1-4, 2007.
- Stalnaker, R. (1979). Assertion. In P. Cole (Ed.), *Syntax and semantics 9: Pragmatics* (p. 315-332). New York: Academic Press.
- Stevenson, R., Crawley, R., & Kleinman, D. (1994). Thematic roles, focusing and the representation of events. *Language and Cognitive Processes*, 9, 519-548.
- Toosarvandani, M. (2014). Contrast and the structure of discourse. *Semantics and*

Pragmatics, 7 (4), 1-57.

- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, 33, 285–318.

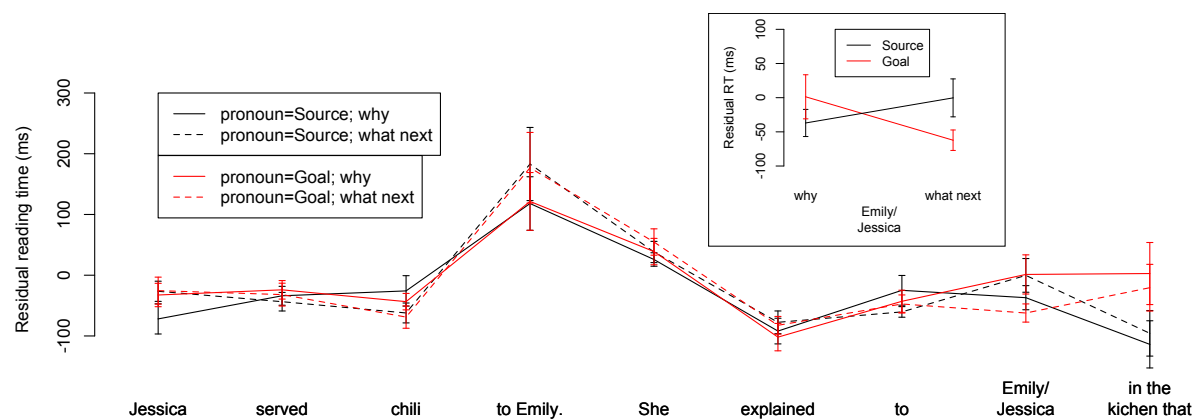


Figure 1. Residual reading times in Experiment 2 (standard errors over participant means). Inset shows interaction at critical region.

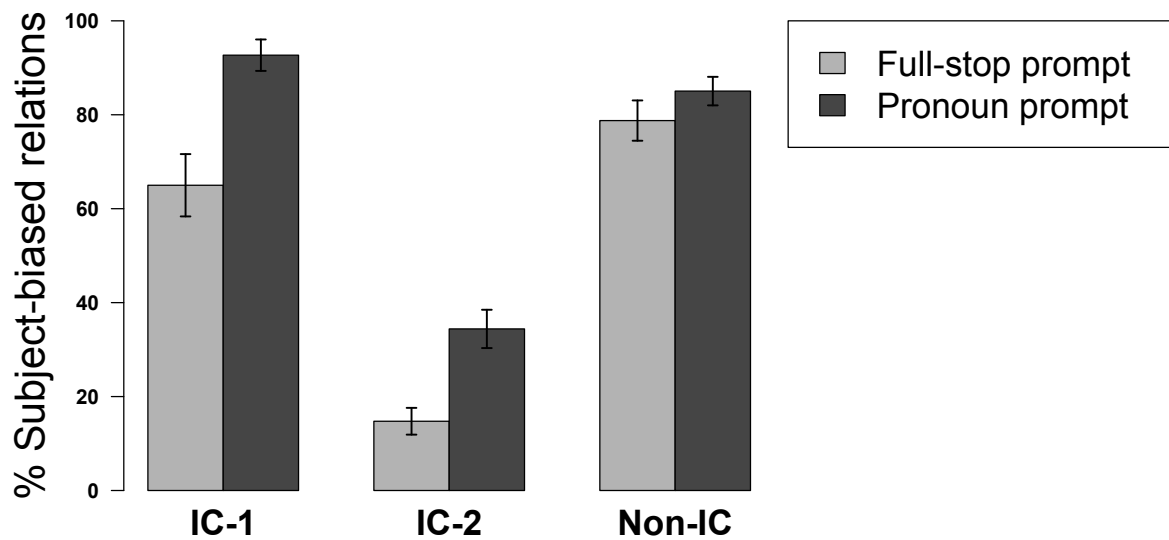


Figure 2. Effect of prompt type on proportions of subject-biased relations by verb class in Experiment 3a (standard errors over participant means)

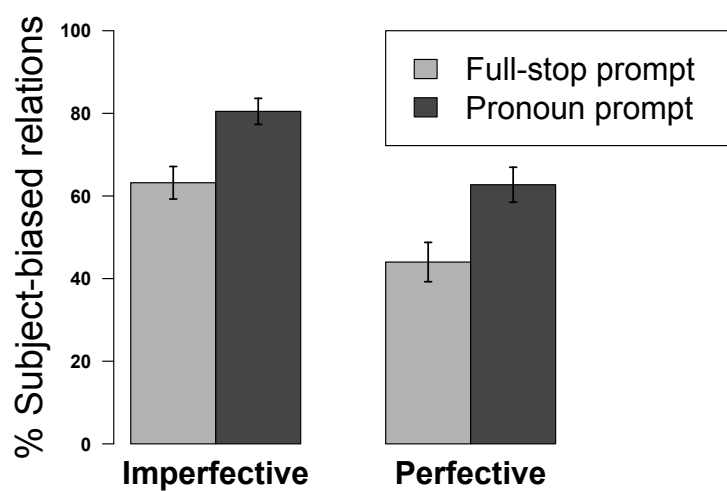


Figure 3. Effect of prompt type on proportions of subject-biased relations by verb aspect in Experiment 3b (standard errors over participant means)

Explanation:	<p>Infer P from the assertion of S_1 and Q from the assertion of S_2, where normally $Q \rightarrow P$.</p> <p><i>We should write a paper. There is a special issue of Discourse Processes coming up that will be dedicated to QUDs. (=1)</i></p>
Result:	<p>Infer P from the assertion of S_1 and Q from the assertion of S_2, where normally $P \rightarrow Q$.</p> <p><i>There is a special issue of Discourse Processes coming up that will be dedicated to QUDs. We should write a paper.</i></p>
Violated Expectation:	<p>Infer P from the assertion of S_1 and Q from the assertion of S_2, where normally $P \rightarrow \neg Q$.</p> <p><i>There is a special issue of Discourse Processes coming up that will be dedicated to QUDs. But I don't think we should write a paper.</i></p>
Occasion:	<p>Infer a change of state for a system of entities from the assertion of S_2, establishing the initial state for this system from the final state of the assertion of S_1.</p> <p><i>Andy walked into his office today. He started working on a paper for the Discourse Processes special issue.</i></p>
Elaboration:	<p>Infer $p(a_1, a_2, \dots)$ from the assertions of S_1 and S_2.</p> <p><i>Andy worked on a paper this evening. He wrote a draft of the introduction and part of the background section.</i></p>
Parallel:	<p>Infer $p(a_1, a_2, \dots)$ from the assertion of S_1 and $p(b_1, b_2, \dots)$ from the assertion of S_2, for a common p and similar a_i and b_i.</p> <p><i>Andy wrote the introduction to a paper this evening. Hannah spent the evening running statistical analyses.</i></p>

Table 1. Definitions and examples of coherence relations

ELABORATION

Friend: Greg corrected Sally.

You: When did this happen?

PARALLEL

Friend: Laura values Luis.

You: Does Luis value Laura?

EXPLANATION

Friend: Ryan hates Amy.

You: What had she done?

RESULT

Friend: James charmed Amber.

You: Did she blush?

OCCASION

Friend: Craig reproached Kate.

You: What happened next?

VIOLATED EXPECTATION

Friend: Ben intimidates Andrea.

You: Why does she like him then?

Table 2. Annotation examples for Experiment 1

	Accuracy
Pronoun=Source; Why	.893 \pm .03
Pronoun=Source; WhatNext	.859 \pm .03
Pronoun=Goal; Why	.867 \pm .02
Pronoun=Goal; WhatNext	.859 \pm .03

Table 3. Question accuracy for Experiment 2 (participant means \pm standard error)

	Name (Raw)	Spill (Raw)	Name (Res)	Spill (Res)
Pronoun=Source; Why	456.5 \pm 33.9	684.0 \pm 65.6	-37.0 \pm 19.9	-113.8 \pm 38.8
Pronoun=Source; WhatNext	456.8 \pm 41.7	620.3 \pm 62.8	-0.3 \pm 27.7	-95.6 \pm 37.5
Pronoun=Goal; Why	519.9 \pm 55.3	789.7 \pm 78.0	1.3 \pm 32.3	2.8 \pm 51.1
Pronoun=Goal; WhatNext	387.3 \pm 26.0	696.9 \pm 72.3	-62.2 \pm 15.0	38.4 \pm 38.4

Table 4. Raw RTs (participant means \pm standard error).